# On the Feasibility of On-body Roaming Models in Human Activity Recognition

Mubarak G. Abdu-Aguye[1], Walid Gomaa[1,2], Yasushi Makihara[3] and Yasushi Yagi[3]

[1] *Computer Science and Engineering Department, Egypt-Japan University of Science and Technology, Egypt*

[2] *Faculty of Engineering, Alexandria University, Egypt*

[3] *The Institute of Scientific and Industrial Research, Osaka University, Japan*

Keywords: Activity Recognition, Transfer Learning, Deep Learning, Data Augmentation.

Abstract: In the domain of human activity recognition, the primary goal is to determine the action a user was performing based on data collected through some sensor modalities. Common modalities adopted to this end include visual and Inertial Measurement Units (IMUs), with the latter taking precedence in recent times due to their unobtrusiveness, low cost and mobility. However, a secondary challenge arises in such sensor-based activity recognition. Difficulties in collecting and annotating training samples are significant and can hinder the performance of models trained on such limited data. As such, there is a need to explore techniques capable of tackling this problem in this domain. In this work, we explore the feasibility of reusing samples collected from different "source" body locations in activity recognition at different "target" body locations. This is achieved through the use of "roaming" models based on recurrent neural networks. We investigate the predictive performance of the transferred samples relative to the performance from samples collected natively at the target body locations. Our results suggest that such roaming models can permit the reuse of cross-body samples without a significant loss in discriminative performance.

## 1 INTRODUCTION

Capturing and analyzing human motion dynamics has garnered interests for several reasons. One potential application is activity recognition, i.e., inference of the physical activities being performed by a person or group of persons within some context, e.g., medical (Liu et al., 2016) or domestic (Hoque and Stankovic, 2012) to name a few. Another application is to enable a robot to mimic human motions for robot-human interaction, as seen in (Elbasiony and Gomaa, 2018).

In order to capture the aforementioned human motion dynamics, different sensing methods can be used, e.g., camera-based (Simonyan and Zisserman, 2014) or sensor-based (Gomaa et al., 2017) methods. While visual (camera-based) methods have seen much adoption due to their efficacy, they are also limited by their lack of mobility and relatively high costs. Sensor-based methods on the other hand, are becoming more common due to the recent ubiquity of sensor-enabled personal devices and their relative unobtrusiveness compared to the other modalities. The sensor-based methods are, however, limited by their locality. That is, most of the sensor-based methods are only able to capture dynamics produced over some small regions of the body (e.g., IMU signals collected at the left elbow can be used only near the left elbow). In order to capture more body-wide/global dynamics, multiple sensors must then be used, each of which is placed at a different body location. This is costly (due to the number of sensors required) and laborious. Additionally, this reduces the feasibility of applications requiring such numbers of sensors due to their lack of user-friendliness.

Specifically in the domain of activity recognition, due to the difficulties inherent in the data collection and annotation process, enabling new human activity recognition systems can require significant cost and effort as described above. In other domains, a shortage of data may be overcome by data augmentation techniques where new samples are generated by some technique with a view to improving the performance of associated machine learning models. However, in this domain data augmentation techniques have not been well studied because of the intuitive difficulties in preserving the semantic meaning of augmented samples.



Abdu-Aguye, M., Gomaa, W., Makihara, Y. and Yagi, Y.
On the Feasibility of On-body Roaming Models in Human Activity Recognition.
DOI: 10.5220/0007921606800690
In *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2019)*, pages 680-690
ISBN: 978-989-758-380-3

In order to overcome such a difficulty, transfer learning can be a potential solution. Transfer Learning (Pan and Yang, 2010) refers to a broad set of techniques which are centered around the reuse of knowledge or data gained in solving some problem in solving some other, related problems. In the former case, the implicit assumption is that there are labelled samples from the unseen problem and what is required is the knowledge of how to use/process these samples accordingly. In the latter case, the implicit assumption is that there is a technique by which samples from some source domain can be re-purposed for use in another domain. As applied here, this has the advantage of mitigating the time and effort required to obtain new samples, which may be significant in domains such as activity recognition. For humanoid robotics, such transfer techniques could also be used to recreate the dynamics of some other body parts without needing to use multiple sensors distributed over different body parts. This would reduce the cost and effort required to design and deploy such systems practically by minimizing the number of sensors required, and also improves the system usability.

In this work we investigate the feasibility of performing data-oriented transfer learning of IMU signals for activity recognition. More specifically, we train a regression model based on long short term memory networks in order to map IMU signals from one body location to another body location. We establish theoretical upper and lower bounds for comparison. We obtain results confirming that the proposed method yielded much better performance than the lower bound, while remaining close to the upper bound due to the transformation process. This indicates the efficacy of the proposed technique.

The rest of this paper is structured as follows. In Section 2, we discuss other literature relevant to this work. In Section 3, we briefly introduce existing techniques used in the proposed method. In Section 4, we describe the details of the proposed method. In Section 5, we give experimental methodology adopted, together with the associated considerations. In Section 6, we analyze the results obtained from our experiments and conclude our work with a summary in Section 7.

## 2 RELATED WORK

In this section we introduce works similar to the proposed method.

(Hu and Yang, 2011) proposed a technique for transfer learning based on sensor mapping. They consider labelled samples from a source domain (i.e set of activities) and aim to perform activity recognition based on unlabelled samples in a target domain (i.e a different set of activities). This is achieved by first building a correspondence between samples from both domains, then performing a label-based transfer between the considered domains. The authors compare it against a baseline unsupervised method and report superior results.

In (Hu et al., 2011), the authors investigated the adaptation of activity instances collected for some task with a view to reusing them for another task. They introduce an algorithm for this purpose, which relies on extant web-accessible metadata describing both sets of activities. From the metadata, a similarity mapping for instances between one activity set and another is derived, together with a confidence estimate for the mapping. Their method was found to yield performance similar to traditionally-proposed techniques.

In (Khan and Roy, 2017), an instance-based transfer learning framework was proposed, where labelled samples from some activity set are re-purposed as samples for a similar activity set, based on some criterion. A combination of clustering and classification is used to build a dynamic pipeline for instance classification between the common and uncommon activity sets. The authors apply their proposed technique to 3 distinct datasets (i.e where each dataset has a different set of activities) and obtained up to an 85% recall in cross-dataset evaluation.



## 3 BACKGROUND

### 3.1 Long Short Term Memory

Long Short Term Memory (LSTM) networks are a type of recurrent neural network (Rumelhart et al., 1986) architecture introduced by (Hochreiter and Schmidhuber, 1997). Traditional recurrent neural networks, while capable of dealing with temporal dependencies in sequential data, have difficulties handling long input sequences.

LSTMs have been shown to yield excellent results on sequential data (e.g (Cho et al., 2014), (Sutskever et al., 2014) etc), often outperforming traditional techniques such as Hidden Markov Models. To handle sequential dependencies, LSTM neurons maintain an internal state which can be considered to "encode" the inputs seen by the neuron up until the current instant of time. By modifying the state, the LSTM cell is capable of adjusting itself to respect the short and long term properties of the input sequence.

The modification of the LSTM's internal state is regulated by structures called Gates. The first of these is called the Forget Gate, which is responsible for 'forgetting'/erasing some portion of the cell's internal state, based on the received input and its previous output. After this operation, the new cell state must be updated based on the received input. This is achieved by the actions of the Input Modulation and Input Gates, the former which generates a candidate internal state value and the latter which specifies how much of the candidate state should be added to the cell's state. The final step in the operation of the LSTM cell is the computation of the actual output value of the cell. This is done by another gate called the Output Gate, which generates an output candidate vector. The cell state is passed through an activation function and the result is multiplied with the output candidate vector to produce the cell's output.

LSTM networks are usually composed of a number of LSTM cells, such that the outputs from preceding cells are used as inputs into proceeding cells and each cell maintains its own cell state. Similar to the traditional recurrent architectures, LSTMs are also trained using Back-Propagation through Time (BPTT) (Werbos et al., 1990).

### 3.2 Bidirectional LSTM

Bidirectional LSTM (Schuster and Paliwal, 1997) is a variant of the LSTM architecture, which is characterized by the parallel stacking of two LSTM networks. The intuition behind the use of bidirectional LSTMs applies specifically when all the time steps of the input data are available at once. In this way the use of such an architecture allows the network to learn both the forward and reverse dynamics of the input, likely capturing more discriminative information than if a single (unidirectional) LSTM was used.

## 4 THE PROPOSED METHOD

### 4.1 Problem Statement

We formalize the problem as follows. We consider two body locations: $s$ denoting the source body location from which the available samples are originally collected, and $t$ denoting the target body location location to which it is desired to transform or roam the samples from $s$ to. Intuitively this can be described as a regression problem. We denote a single, multidimensional finite-length sample from $s$ as $d_s$ and a single, multidimensional finite-length sample from $t$ as $d_t$. Note that $d_t$ and $d_s$ are collected over the same time window. We can then formulate the problem mathematically as:

$$d_t = F(d_s), \tag{1}$$

where $F()$ is a roaming function from the source to the target.

From Eq. (1) it can be seen that the roaming problem can be expressed as the learning of a suitable roaming function $F()$ such that the difference between $F(d_s)$ and $d_t$ (since they are collected at the same instants of time) is minimized. Since a neural network can be considered to be a universal function approximator (Cybenko, 1989), we consider the learning of an approximation $\hat{F}()$ of the function $F()$ from collected data samples using a deep neural network with a view to obtaining this optimal mapping.

### 4.2 Network Architecture

We carry out the following procedure to construct the roaming models described previously. In this work, we consider only the tri-axial accelerometer and tri-axial gyroscope sensors as we believe that these two sensors capture the most relevant information about the dynamics of the signals under consideration. We segment the source and target data samples into 1 second windows based on the work of (Banos et al., 2014a).

We consider an arbitrary pair of body locations as source ($s$) and target ($t$) locations. Next, we construct a deep neural network to approximate the roaming function as described previously. Although we considered different architectures (fully-connected, convolutional, etc) we found that recurrent architectures obtained the best performance. This can be attributed to their inherent affinity for sequential input data. Therefore we construct a deep recurrent neural network as shown in Figure 1. It consists of two bidirectional LSTM layers, each containing 100 LSTM cells. We treat the first LSTM layer as an encoder and the second as a decoder, such that the dynamics of the source samples are first encoded into the cell states of the encoder LSTM. Subsequently, the decoder LSTM aims to use the captured dynamics as input and reconstruct the corresponding target location dynamics. The final output of the decoder LSTM layer is fed through a single fully-connected layer consisting of six neurons.

During model training, sample windows (of 1s length as stated above) and fed as the input and desired output of the network. We train the network (i.e the two LSTMs in tandem) using adaptive moment estimation (Kingma and Ba, 2015) as the weight optimization technique. Since this is a regression problem, we use the mean squared error as the loss func-

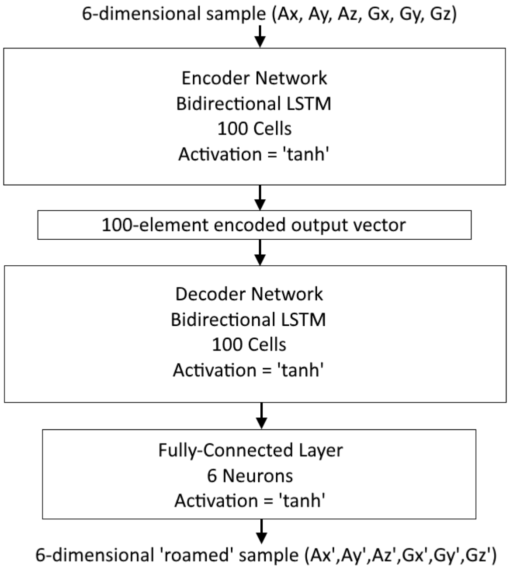


Figure 1: Architecture of Roaming Model.

tion so as to obtain an optimal reconstruction of the target samples from the source samples.

We consider different sets of body locations. Due to space constraints we limit our analyses to three body locations in the upper half of the body - the right lower arm (RLA), right upper arm (RUA) and the back (BACK) - and construct roaming models between each pair of locations.

# 5 EXPERIMENTS

## 5.1 Dataset

We used the REALDISP Dataset to evaluate the proposed method. The REALDISP dataset (Baños et al., 2012), (Banos et al., 2014b) is an activity recognition dataset consisting of multimodal sensor readings (accelerometer, gyroscope, and magnetometer) obtained using Xsens IMU units at a sample rate of 50Hz. It contains samples collected from three different device placements: one of which is considered ideal, and the others with increasing positional displacements. We use the data corresponding to the ideal placement in this work. 17 subjects were involved in the collection of the data, and there are 33 considered activities spanning upper extremity, lower extremity and full-body activity sets. The constituent activities are shown in Table 1.

During the data capture process multiple sensor units were attached to the subjects' bodies. Specifically, 9 body locations were considered which are

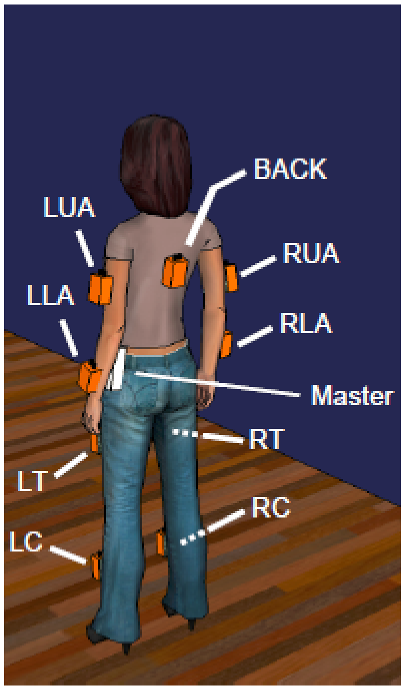


Figure 2: Device Placements in REALDISP dataset. Image reused with permission from (Baños et al., 2012), (Banos et al., 2014b).

shown in Figure 2. Therefore for any given point in time $t$, the collected data contains multimodal samples from 9 body locations all collected at that instant $t$. After we segment the raw samples into 1 second long windows, we obtain a total of 13963 samples. This makes it possible to build a roaming model between on-body locations as proposed in this work.

## 5.2 Evaluation Methodology

We evaluate the quality of the roamed samples and by extension, the roaming models as follows. For any source location $s$ and target location $t$, we segment the available data into the following proportions - 50%, 25%, 25%. We denote the different portions of data as described in Figure 3.

We carry out three sets of evaluations for all the roaming scenarios considered; the first of these is aimed at determining a theoretical lower bound for predictive performance, and this is done by training a classifier on 25% of the samples from the target location (1B in Figure 3), then testing 25% of the samples from the source location (0C in Figure 3) using this classifier. This is done to obtain an estimate of the predictive performance obtainable *without* the use of the proposed roaming method. In the second evaluation, we train a roaming model on the 50% portions of the source and target data (denoted as 0A and 1A in

Table 1: Activities in REALDISP dataset.

| | | |
|---|---|---|
| Walking | Waist Rotation | Shoulders High Amplitude Rotation |
| Jogging | Waist Bends | Shoulders Low Amplitude Rotation |
| Running | Reach Heels Backwards | Arms Inner Rotation |
| Jump Up | Lateral Bend | Knees (alternately) to breast |
| Jump Front& Back | Lateral Bend Arm Up | Heels (alternately) to backside |
| Jump Sideways | Repetitive forward stretching | Knees Bending (crouching) |
| Jump leg/arms open/closed | Upper Trunk & Lower Body Opposite Twist | Knees (alternately) bend forward |
| Jump Rope | Arms Lateral Elevation | Rotation on the knees |
| Trunk twist (arms outstretched) | Arms frontal elevation | Rowing |
| Trunk twist (elbows bent) | Frontal hand claps | Elliptic Bike |
| Waist bends forward | Arms frontal crossing | Cycling |

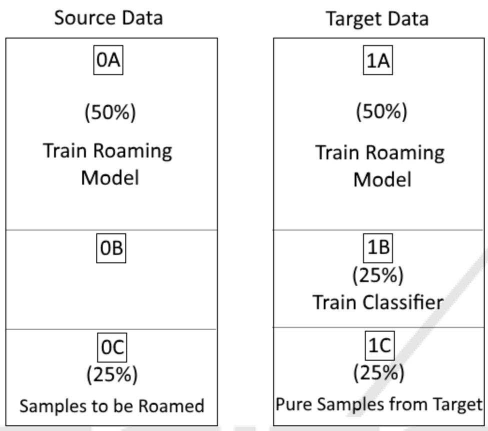


Figure 3: Data assignment for experimental and evaluation methodology.

Figure 3), then roam 25% of the source samples (0C in Figure 3) using the trained model. The classifier from the first evaluation is then used to evaluate the roamed samples i.e the roamed samples are used as the testing data. This is done with a view to deriving the performance gains obtained through the use of the roaming model. Finally, we once again use the trained classifier, but evaluate on samples collected natively from the target body location (1C in Figure 3). This is done in order to obtain a theoretical upper bound for the proposed method, as well as give a quantitative value for the loss in predictive performance incurred through the use of the roaming model (i.e the loss incurred from re-purposing samples rather than collecting them anew at the target location).

We consider all the activities in the REALDISP dataset, as we aim to investigate the feasibility of transferring samples in general between body locations, independent of any particular activity. Because of the wide variety of activities in this dataset, we feel that the performance obtained from our experiments will be a reasonable indicator of the feasibility of such roaming models in general. All the considered evaluations are carried out across-subjects i.e the subjects in *A, *B and *C are all different. This is done to

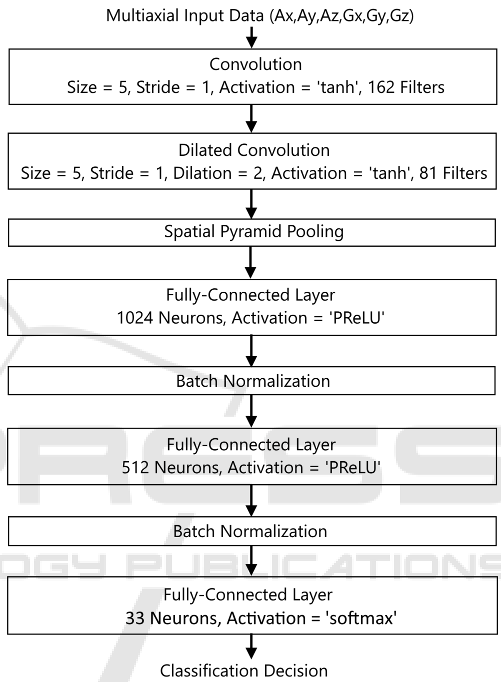


Figure 4: Structure of CNN used for classification.

simulate practical scenarios in the real world where model training and inference are likely done on disjoint sets of individuals. We use a Convolutional Neural Network (CNN) as the classifier due to its ability to automatically extract features from the source data. The structure of the adopted CNN is shown in Figure 4.

The described evaluations (baseline and roaming/comparative) are repeated ten times and averaged in order to obtain a general estimate of the quality of the roamed samples. For every evaluation, Different subject selections are made for training and testing purposes. The Keras machine learning library was used for the implementation of the roaming model, while the classifier was implemented using the PyTorch machine learning library. Due to space limitations, we consider the average classification accuracy and F-metric as the evaluation metrics. The obtained results and their discussion are provided in the following section.

# 6 RESULTS

In this section we present the results of our experiments and analyze them further. In the result tables we indicate activities involving the whole body with ‡, the trunk with ||, the upper extremities with † and the lower extremities with $\wedge$. Additionally we group the rows per activity subgroup (i.e whole body, upper extremities, etc) for easier interpretability. The theoretical lower bound results (i.e from training on the target location and using samples from the source location directly) are presented under the column marked "LB". The results obtained from the proposed method (i.e roaming samples from the source to the target location and using the classifier trained on the target location) are listed under the "Proposed" column. The results from the theoretical upper bound (i.e training and testing on samples from the target location) are shown under the "UB" column. All the F-metrics are reported on a normalized scale i.e 0 to 1.

## 6.1 Right Lower Arm to Right Upper Arm and Back

Table 2 shows the results of the roaming experiments in which the right lower arm (RLA) is the source location and the right upper arm (RUA) and back (BACK) are the target locations.

For the activities involving the whole body, trunk and upper extremities, the right upper arm (RUA) arm upper bound shows good F-metrics, as compared to the activities involving the lower extremities. This is because the preceding sets of activities have more dynamics captured by the right upper arm as compared to the latter set of activities. Additionally the lower bound results indicate that the RLA generates roughly similar dynamics as generated by the right upper arm due to their relative proximity. As such, the results obtained by the proposed method show significant improvement over the lower bound, indicating that the roaming model is effective in this scenario.

When the back (BACK) is considered as the target location, it can be seen that the lower bounds are significantly worse than those obtained from the right upper arm, with several activities completely mis-classified (most clearly seen with the activities involving the upper extremities). This is due to the fact that the classifier is unable to recover the dynamics of the back from the right lower arm without transformation. After the transformation obtained through the use of the proposed method, the predictive performance obtained per activity improves over the lower bound for all activities. The least performance benefit obtained is correspondingly observed for the activities involving the upper extremities, indicating that roaming from the right lower arm to the back is most effective for activities of the whole body, trunk or lower extremities.

## 6.2 Right Upper Arm to Right Lower Arm and Back

Table 3 shows the results obtained using our proposed technique when the right upper arm (RUA) is the source location and the right lower arm (RLA) and back (BACK) are the target locations.

From the upper bound results for the former scenario (i.e RUA-RLA), it can be observed that the RLA location shows fairly good recognition over all activity subsets (whole body, trunk, upper extremities, etc), with the poorest performance observed for the lower extremity activities. Similar to the previous experiments, this can be attributed to the fact that the RLA location does not capture most of the dynamics generated by such activities. The lower bound results for same are expectedly much lower, but show an aggregate predictive accuracy similar to the converse case as seen in Table 2 i.e RLA-RUA yields roughly similar performance to RUA-RLA. This can be explained as before; the proximity of the two locations permits them to be affected broadly similarly by the same activities. After the application of the proposed method, a significant improvement is observed in the predictive performance for all the considered activities. The activities involving the upper extremities show the best roaming performance, in some cases showing virtually no degradation in performance relative to the upper bound. This can similarly be attributed to the proximity of the source and target location, which permits them to be subjected to similar dynamics and therefore yield similar predictive performance.

Considering the latter scenario (i.e RUA-BACK), it can be seen (from the upper bound results) that the BACK location shows good recognition accuracy for all activity subsets except for the upper extremity activities where its performance degrades significantly. This can be attributed to the fact that the BACK location is very minimally affected by movements of the arm and therefore cannot be used to discriminate between such activities accurately. The lower bound results also suggest that samples from the RUA location

Table 2: F-measures for Roaming Experiments for Right Lower Arm (RLA) to Right Upper Arm (RUA) and Back (BACK).

| **Source Location** | RLA | | | | | |
|---|---|---|---|---|---|---|
| **Target Location** | RUA | | | BACK | | |
| **Method** | **LB** | **Proposed** | **UB** | **LB** | **Proposed** | **UB** |
| **Average Classification Accuracy (%):** | **21.21** | **66.05** | **73.48** | **4.66** | **44.18** | **63.35** |
| Walking‡ | 0.55 | 0.71 | 0.83 | 0.00 | 0.63 | 0.78 |
| Jogging‡ | 0.04 | 0.72 | 0.77 | 0.00 | 0.54 | 0.70 |
| Running‡ | 0.02 | 0.70 | 0.79 | 0.01 | 0.63 | 0.85 |
| Jump Up‡ | 0.28 | 0.52 | 0.59 | 0.02 | 0.38 | 0.40 |
| Jump Front & Back‡ | 0.36 | 0.56 | 0.64 | 0.16 | 0.41 | 0.68 |
| Jump Sideways‡ | 0.11 | 0.41 | 0.50 | 0.11 | 0.44 | 0.65 |
| Jump legs/arms open/closed‡ | 0.21 | 0.86 | 0.92 | 0.04 | 0.35 | 0.54 |
| Jump Rope‡ | 0.24 | 0.57 | 0.60 | 0.03 | 0.24 | 0.35 |
| Rowing‡ | 0.07 | 0.64 | 0.66 | 0.01 | 0.55 | 0.64 |
| Elliptic Bike‡ | 0.00 | 0.71 | 0.63 | 0.00 | 0.46 | 0.56 |
| Cycling‡ | 0.05 | 0.57 | 0.78 | 0.00 | 0.11 | 0.47 |
| Trunk Twist (arms out)‖ | 0.41 | 0.83 | 0.85 | 0.02 | 0.55 | 0.59 |
| Trunk Twist (elbows bent)‖ | 0.00 | 0.94 | 0.94 | 0.04 | 0.51 | 0.63 |
| Waist Bent Forward‖ | 0.16 | 0.56 | 0.75 | 0.01 | 0.69 | 0.92 |
| Waist Rotation‖ | 0.02 | 0.65 | 0.76 | 0.05 | 0.45 | 0.77 |
| Waist Bends (reach foot w/ opposite hand)‖ | 0.27 | 0.78 | 0.92 | 0.00 | 0.67 | 0.87 |
| Reach Heel Backwards‖ | 0.46 | 0.52 | 0.67 | 0.20 | 0.29 | 0.56 |
| Lateral Bend‖ | 0.14 | 0.51 | 0.54 | 0.11 | 0.42 | 0.62 |
| Lateral Bend Arm Up‖ | 0.17 | 0.41 | 0.49 | 0.05 | 0.32 | 0.57 |
| Repetitive Forward Stretching‖ | 0.15 | 0.68 | 0.79 | 0.00 | 0.63 | 0.90 |
| Upper Trunk & Lower Body Opposite Twist‖ | 0.13 | 0.42 | 0.61 | 0.01 | 0.10 | 0.51 |
| Arms Lateral Elevation† | 0.12 | 0.70 | 0.59 | 0.00 | 0.10 | 0.23 |
| Arms Frontal Elevation† | 0.07 | 0.69 | 0.72 | 0.00 | 0.14 | 0.25 |
| Frontal Hand Claps† | 0.38 | 0.68 | 0.78 | 0.00 | 0.25 | 0.40 |
| Arms Frontal Crossing† | 0.29 | 0.83 | 0.84 | 0.00 | 0.17 | 0.49 |
| Shoulders High-Amplitude Rotation† | 0.72 | 0.89 | 0.91 | 0.01 | 0.13 | 0.33 |
| Shoulders Low-Amplitude Rotation† | 0.22 | 0.87 | 0.83 | 0.01 | 0.23 | 0.29 |
| Arms Inner Rotation† | 0.06 | 0.76 | 0.72 | 0.00 | 0.11 | 0.60 |
| Knees (alternately) to breast∧ | 0.17 | 0.71 | 0.86 | 0.14 | 0.63 | 0.88 |
| Heels (alternately) to backside∧ | 0.35 | 0.73 | 0.73 | 0.23 | 0.59 | 0.80 |
| Knees bending (crouching)∧ | 0.31 | 0.31 | 0.35 | 0.24 | 0.32 | 0.60 |
| Knees (alternately) bent forward∧ | 0.18 | 0.43 | 0.59 | 0.02 | 0.28 | 0.54 |
| Rotation on Knees∧ | 0.01 | 0.69 | 0.69 | 0.00 | 0.56 | 0.74 |



show very different dynamics than those collected natively from the BACK location and as such show extremely poor recognition performance in that scenario (i.e without roaming). After roaming, the predictive performance improves drastically across all activities, with the aggregate performance (as measured by the average accuracy) approaching the aggregate performance of the native BACK-collected samples. The worst roaming performance is observed for activities involving the upper extremities, indicating that roaming from the RUA location to the BACK location is not effective for upper extremity activities.

### 6.3 Back to Right Upper Arm and Right Lower Arm

Table 4 shows the results obtained using the proposed method when the back (BACK) is considered as the source location and the right lower arm (RLA) and right upper arm (RUA) are considered as the target locations.

It can be observed that the RLA location natively (i.e from the upper bound results) shows the best predictive performance for activities involving the upper extremities. This can be explained by the fact

Table 3: F-measures for Roaming Experiments for Right Upper Arm (RUA) to Right Lower Arm (RLA) and Back (BACK).

| **Source Location** | RUA | | | | | |
|---|---|---|---|---|---|---|
| **Target Location** | RLA | | | BACK | | |
| **Method** | **LB** | **Proposed** | **UB** | **LB** | **Proposed** | **UB** |
| **Average Classification Accuracy (%):** | **23.01** | **64.13** | **71.21** | **9.02** | **60.18** | **66.32** |
| Walking[‡] | 0.43 | 0.68 | 0.79 | 0.08 | 0.80 | 0.76 |
| Jogging[‡] | 0.00 | 0.61 | 0.66 | 0.03 | 0.67 | 0.73 |
| Running[‡] | 0.00 | 0.70 | 0.69 | 0.30 | 0.72 | 0.80 |
| Jump Up[‡] | 0.27 | 0.46 | 0.48 | 0.00 | 0.43 | 0.47 |
| Jump Front & Back[‡] | 0.30 | 0.50 | 0.51 | 0.20 | 0.60 | 0.68 |
| Jump Sideways[‡] | 0.24 | 0.40 | 0.45 | 0.02 | 0.47 | 0.61 |
| Jump legs/arms open/closed[‡] | 0.13 | 0.85 | 0.84 | 0.12 | 0.52 | 0.61 |
| Jump Rope[‡] | 0.29 | 0.81 | 0.85 | 0.02 | 0.31 | 0.48 |
| Rowing[‡] | 0.04 | 0.56 | 0.73 | 0.01 | 0.63 | 0.78 |
| Elliptic Bike[‡] | 0.00 | 0.72 | 0.82 | 0.06 | 0.55 | 0.63 |
| Cycling[‡] | 0.15 | 0.60 | 0.63 | 0.00 | 0.68 | 0.66 |
| Trunk Twist (arms out)[‖] | 0.28 | 0.77 | 0.88 | 0.00 | 0.60 | 0.56 |
| Trunk Twist (elbows bent)[‖] | 0.01 | 0.83 | 0.87 | 0.20 | 0.70 | 0.64 |
| Waist Bent Forward[‖] | 0.26 | 0.60 | 0.61 | 0.00 | 0.85 | 0.89 |
| Waist Rotation[‖] | 0.04 | 0.49 | 0.70 | 0.01 | 0.69 | 0.80 |
| Waist Bends (reach foot w/ opposite hand)[‖] | 0.44 | 0.75 | 0.80 | 0.12 | 0.86 | 0.83 |
| Reach Heel Backwards[‖] | 0.37 | 0.48 | 0.54 | 0.07 | 0.44 | 0.64 |
| Lateral Bend[‖] | 0.24 | 0.54 | 0.58 | 0.02 | 0.63 | 0.74 |
| Lateral Bend Arm Up[‖] | 0.26 | 0.47 | 0.56 | 0.01 | 0.55 | 0.67 |
| Repetitive Forward Stretching[‖] | 0.18 | 0.62 | 0.55 | 0.00 | 0.78 | 0.81 |
| Upper Trunk & Lower Body Opposite Twist[‖] | 0.12 | 0.34 | 0.48 | 0.05 | 0.17 | 0.48 |
| Arms Lateral Elevation[†] | 0.25 | 0.61 | 0.76 | 0.00 | 0.22 | 0.23 |
| Arms Frontal Elevation[†] | 0.16 | 0.68 | 0.83 | 0.00 | 0.33 | 0.37 |
| Frontal Hand Claps[†] | 0.39 | 0.70 | 0.85 | 0.00 | 0.43 | 0.52 |
| Arms Frontal Crossing[†] | 0.15 | 0.85 | 0.86 | 0.01 | 0.27 | 0.45 |
| Shoulders High-Amplitude Rotation[†] | 0.81 | 0.93 | 0.96 | 0.00 | 0.32 | 0.38 |
| Shoulders Low-Amplitude Rotation[†] | 0.15 | 0.88 | 0.89 | 0.00 | 0.27 | 0.43 |
| Arms Inner Rotation[†] | 0.00 | 0.68 | 0.84 | 0.00 | 0.49 | 0.55 |
| Knees (alternately) to breast[∧] | 0.31 | 0.60 | 0.45 | 0.23 | 0.87 | 0.94 |
| Heels (alternately) to backside[∧] | 0.25 | 0.56 | 0.56 | 0.12 | 0.77 | 0.86 |
| Knees bending (crouching)[∧] | 0.25 | 0.32 | 0.29 | 0.15 | 0.56 | 0.67 |
| Knees (alternately) bent forward[∧] | 0.12 | 0.41 | 0.61 | 0.20 | 0.49 | 0.64 |
| Rotation on Knees[∧] | 0.00 | 0.65 | 0.82 | 0.00 | 0.62 | 0.70 |

that the effects of such activities are localized mostly along the aforementioned extremities, the data from which were used in training the classifier in that experiment. Conversely, it shows the worst performance for activities involving the lower extremities. This can be attributed to the fact that such activities do not generate appreciable effects in the upper extremities. The lower bound results also show that the dynamics of the back differ significantly from the dynamics of the RLA location. This is reflected in the poor predictive performance obtained, with several activities completely mis-classified. After the samples are roamed, improvements in the predictive performance are apparent; this is most apparent with the whole-body activities and the activities involving the trunk. Accordingly, the improvements are less in the activities involving the upper extremities. This can be explained by the fact that the BACK location natively does not capture much of the dynamics of the upper extremities, and as such this limits the utility of roaming samples from that location to the upper extremities.

Accordingly, the RUA location natively (i.e from the upper bound results) shows the best predictive performance of all the considered target locations, showing good performance across all the activities al-

Table 4: F-measures for Roaming Experiments for Back (BACK) to Right Lower Arm (RLA) and Right Upper Arm (RUA).

| **Source Location** | BACK | | | | | |
|---|---|---|---|---|---|---|
| **Target Location** | RLA | | | RUA | | |
| **Method** | **LB** | **Proposed** | **UB** | **LB** | **Proposed** | **UB** |
| **Average Classification Accuracy (%):** | **7.16** | **43.80** | **67.19** | **11.63** | **56.82** | **78.03** |
| Walking‡ | 0.03 | 0.54 | 0.74 | 0.11 | 0.70 | 0.88 |
| Jogging‡ | 0.00 | 0.49 | 0.53 | 0.08 | 0.61 | 0.74 |
| Running‡ | 0.00 | 0.67 | 0.64 | 0.56 | 0.59 | 0.77 |
| Jump Up‡ | 0.18 | 0.26 | 0.46 | 0.20 | 0.41 | 0.61 |
| Jump Front & Back‡ | 0.29 | 0.41 | 0.49 | 0.26 | 0.55 | 0.66 |
| Jump Sideways‡ | 0.07 | 0.36 | 0.42 | 0.18 | 0.48 | 0.60 |
| Jump legs/arms open/closed‡ | 0.01 | 0.58 | 0.73 | 0.12 | 0.75 | 0.92 |
| Jump Rope‡ | 0.02 | 0.41 | 0.82 | 0.07 | 0.40 | 0.70 |
| Rowing‡ | 0.00 | 0.44 | 0.71 | 0.01 | 0.44 | 0.72 |
| Elliptic Bike‡ | 0.00 | 0.32 | 0.71 | 0.00 | 0.52 | 0.77 |
| Cycling‡ | 0.00 | 0.28 | 0.54 | 0.00 | 0.51 | 0.85 |
| Trunk Twist (arms out)‖ | 0.00 | 0.55 | 0.82 | 0.01 | 0.66 | 0.86 |
| Trunk Twist (elbows bent)‖ | 0.00 | 0.60 | 0.91 | 0.09 | 0.75 | 0.97 |
| Waist Bent Forward‖ | 0.20 | 0.52 | 0.50 | 0.00 | 0.81 | 0.84 |
| Waist Rotation‖ | 0.12 | 0.38 | 0.72 | 0.00 | 0.62 | 0.78 |
| Waist Bends (reach foot w/ opposite hand)‖ | 0.00 | 0.65 | 0.71 | 0.18 | 0.91 | 0.93 |
| Reach Heel Backwards‖ | 0.13 | 0.22 | 0.59 | 0.13 | 0.47 | 0.65 |
| Lateral Bend‖ | 0.16 | 0.42 | 0.60 | 0.01 | 0.35 | 0.64 |
| Lateral Bend Arm Up‖ | 0.06 | 0.21 | 0.53 | 0.06 | 0.44 | 0.59 |
| Repetitive Forward Stretching‖ | 0.00 | 0.48 | 0.53 | 0.00 | 0.76 | 0.82 |
| Upper Trunk & Lower Body Opposite Twist‖ | 0.00 | 0.10 | 0.59 | 0.01 | 0.18 | 0.64 |
| Arms Lateral Elevation† | 0.00 | 0.13 | 0.77 | 0.00 | 0.09 | 0.71 |
| Arms Frontal Elevation† | 0.00 | 0.27 | 0.83 | 0.01 | 0.35 | 0.87 |
| Frontal Hand Claps† | 0.00 | 0.55 | 0.82 | 0.00 | 0.47 | 0.85 |
| Arms Frontal Crossing† | 0.00 | 0.41 | 0.84 | 0.00 | 0.26 | 0.90 |
| Shoulders High-Amplitude Rotation† | 0.00 | 0.33 | 0.95 | 0.00 | 0.42 | 0.96 |
| Shoulders Low-Amplitude Rotation† | 0.00 | 0.43 | 0.93 | 0.00 | 0.45 | 0.90 |
| Arms Inner Rotation† | 0.00 | 0.49 | 0.86 | 0.00 | 0.53 | 0.91 |
| Knees (alternately) to breast∧ | 0.32 | 0.54 | 0.47 | 0.32 | 0.79 | 0.82 |
| Heels (alternately) to backside∧ | 0.19 | 0.62 | 0.66 | 0.23 | 0.78 | 0.79 |
| Knees bending (crouching)∧ | 0.16 | 0.27 | 0.27 | 0.25 | 0.47 | 0.47 |
| Knees (alternately) bent forward∧ | 0.01 | 0.30 | 0.61 | 0.22 | 0.43 | 0.57 |
| Rotation on Knees∧ | 0.00 | 0.48 | 0.68 | 0.01 | 0.71 | 0.76 |



though the upper extremity activities show the best performance. Similar to the RLA location, the predictive performance obtained for lower-body extremities is generally the lowest obtained, but the RUA location shows better robustness than the RLA location. This is due to the fact that the RUA location, being higher-up along the body, is able to capture some of the dynamics of the lower extremity activities, as opposed to the RLA location which is more particular/local to the upper extremities. Before the proposed technique is applied, it can be seen that the direct reuse of the BACK location samples at the RUA location yields poor predictive performance. This can be explained similarly to the previous section i.e the dynamics of the two locations are differ significantly. Regardless, the difference between the BACK and RUA locations' dynamics is lower than the difference between the BACK and RLA locations. This can be seen in the lower number of completely mis-classified samples for the BACK-RUA scenario compared to the BACK-RLA scenario. When the proposed method is applied, improvements in the predictive performance are obtained for all considered activities. The best roaming performance gains are observed in activities involving the lower extremities, followed by the activities involving the trunk. This can be attributed to the fact that the BACK location captures more of the dynamics of the lower extremities than the RUA location. Therefore by roaming the samples from the BACK location to the RUA location, together with the relative similarity in the dynamics captured by both locations, near-native predictive performance is gained on activities whose effects have an impact on both the BACK and RUA locations.

### 6.4 Discussion

In this section we summarize inferences obtained from the results and their subsequent discussion in the preceding subsections.

Considering the general pattern of results obtained, it can be inferred that the proposed roaming technique performs best between proximal body locations (e.g RLA-RUA, RUA-BACK, etc). This can be attributed to the fact that proximal body locations are similarly affected by the dynamics of a given set of activities. Additionally, roaming performance is maximized when the source body location is itself suitably discriminative of the considered activities. This follows due to the reason that the source location must itself contain sufficient recognition information, which may then be roamed for reuse. We can thus surmise that, in general, the predictive performances obtained through roaming (relative to the reported upper-bounds) indicate that the roamed samples can be useful in solving the activity recognition problem.

## 7 CONCLUSION AND FUTURE WORK

In this work we explore the feasibility of constructing "roaming" models, which are capable of transforming samples between different body locations. We propose a technique based on Bidirectional Recurrent Neural Networks and apply our proposed technique to a multi-sensor, multi-location dataset consisting of 33 activities involving different regions of the body. The results obtained suggest that roaming models can be feasible between proximal body locations, and the highest gains from their use result from the consideration of activities whose dynamics are adequately captured by the source location.

In the future we would like to explore more body locations to gain a more holistic insight into the performance and feasibility of such models. Additionally, we will investigate roaming models based on the use of more than one source body location. Furthermore, the feasibility of activity class-specific (e.g upper body activities) roaming models may be explored in order to obtain better roaming performance as opposed to using a general roaming model for all activities.

## REFERENCES

Baños, O., Damas, M., Pomares, H., Rojas, I., Tóth, M. A., and Amft, O. (2012). A benchmark dataset to evaluate sensor displacement in activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 1026–1035. ACM.

Banos, O., Galvez, J.-M., Damas, M., Pomares, H., and Rojas, I. (2014a). Window size impact in human activity recognition. *Sensors*, 14(4):6474–6499.

Banos, O., Toth, M., Damas, M., Pomares, H., and Rojas, I. (2014b). Dealing with the effects of sensor displacement in wearable activity recognition. *Sensors*, 14(6):9995–10023.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.

Elbasiony, R. and Gomaa, W. (2018). Humanoids skill learning based on real-time human motion imitation using kinect. *Intelligent Service Robotics*, 11(2):149–169.

Gomaa, W., Elbasiony, R., and Ashry, S. (2017). Adl classification based on autocorrelation function of inertial signals. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 833–837. IEEE.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hoque, E. and Stankovic, J. (2012). Aalo: Activity recognition in smart homes using active learning in the presence of overlapped activities. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 139–146. IEEE.

Hu, D. H. and Yang, Q. (2011). Transfer learning for activity recognition via sensor mapping. In *Twenty-second international joint conference on artificial intelligence*.

Hu, D. H., Zheng, V. W., and Yang, Q. (2011). Cross-domain activity recognition via transfer learning. *Pervasive and Mobile Computing*, 7(3):344–358.

Khan, M. A. A. H. and Roy, N. (2017). Transact: Transfer learning enabled activity recognition. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 545–550.

Kingma, D. and Ba, J. (2015). Adam: a method for stochastic optimization (2014). *arXiv preprint arXiv:1412.6980*, 15.

Liu, X., Liu, L., Simske, S. J., and Liu, J. (2016). Human daily activity recognition for healthcare using wearable and visual sensing data. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 24–31. IEEE.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propogating errors, nature 323, 533-536.

Schuster, M. and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.

Simonyan, K. and Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Werbos, P. J. et al. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560.